Synopsis for Registration as Research Student

For

Doctoral Degree

On

Content Based Bio-Medical Image Retrieval System

By

Galshetwar Gajanan Madhavrao

Research Scholar

Supervisor

Dr. L. M. Waghmare

Department of Electronics and Telecommunication Engineering,

Shri Guru Gobind Singhji Institute of Engineering and Technology,

Nanded-431606. (MS.) India.

1. Introduction

In last few years there has been a rapid growth in severe and critical diseases in India and all over the world. As a result of this growth there is increasing need of expert medical services in urban as well as in remote places especially in developing countries. In areas where general clinical practices are present we can provide them a technological solution which can assist those clinics bring expert medical services to their help. There have been many technological developments in healthcare areas which helped us to take expert medical services to remote places of world. Telemedicine is one of the significant developments in recent technology that has enabled medical science to reach to the patients door step. Other technology that enabled us to provide engineering solutions to meet these increasing needs of expert services is Bio-medical imaging. Bio-medical imaging is used to create visual representation of different interior parts of body which are used for medical analysis and diagnosis. Nowadays, these bio-medical images recorded from patients are being stored to create large databases which can be later used for different purposes. We engineers can use such databases for designing autonomous systems which will help doctors in analysing the patients' condition. There are several types of bio-medical images few of which that are popularly known are as follows:

1.1 Fundus Images

Fundus Images are taken mainly to analyse a critical disease known as Diabetic retinopathy (DR). Diabetic patients are generally vulnerable to DR which is associated with the retina of the patients and make them prone to complete blindness if untreated in time. Diabetic patients should go for a regular check-up and detect DR at initial stage only. DR grows step by step and reaches to a very severe stage from where it is nearly impossible to cure the retina.

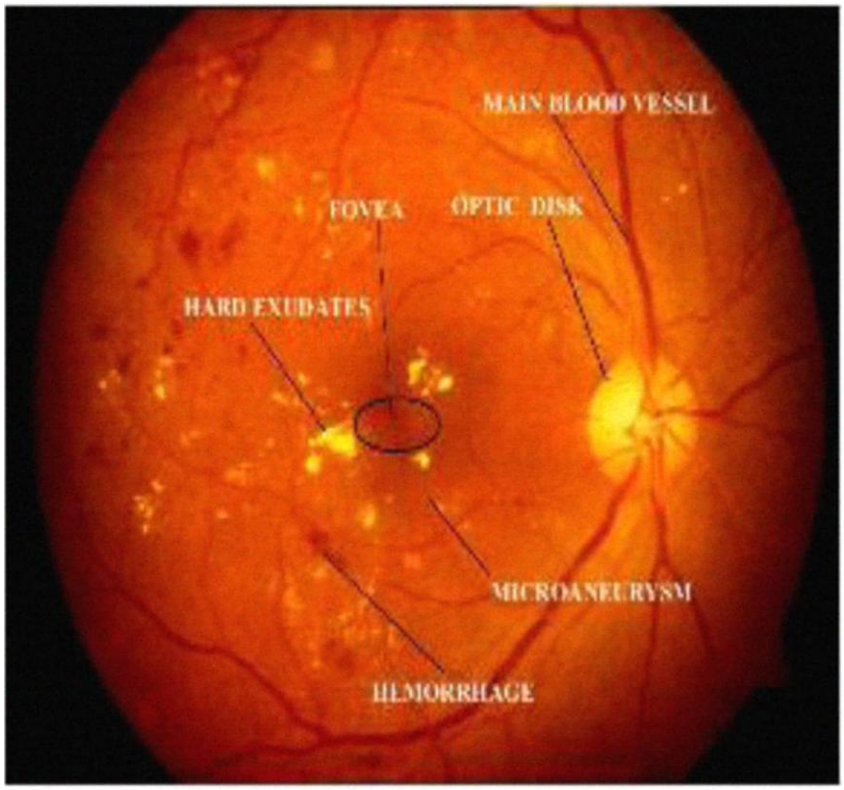

Figure 1: Fundus image showing symptoms of Diabetic retinopathy.

The different stages of DR are as follows:

a. Micro-aneurysms: - This is the very first detectable stage of DR which results when there is change in the size of retinal capillaries.
b. Hemorrhages:- This is also a primary stage of DR which arises after Micro-aneurysms appear in retina. This results due to ruptured Micro-aneurysms.
c. Hard exudates: - These result due to leakages from the blood vessels in retina.
d. Soft exudates: - These are the later stage of DR where the lipids from vessels is obstructed which forms Microinfarcts. These microinfarcts are also known as soft exudates.
e. Neovascularization: - The soft exudates cause lack of oxygen which in turn causes new fragile vessels. This stage is known as neovascularization. This can cause sudden loss of sight or permanent blindness.

These symptoms are more serious if located near the centre of the macula. To prevent all this, early detection of DR is must and to do so regular check-up of diabetic patients is necessary. DR is treatable disease but can be fatal if ignored.

1.2 Computerized Tomography (CT) Images

Computed Tomography (CT), also commonly referred to as a Computerized axial tomography (CAT) scan, is a medical imaging method that combines multiple X-ray projections taken from different angles to produce detailed cross-sectional images of areas inside the body. CT images allow doctors to get very precise, 3-D views of certain parts of the body, such as soft tissues, the pelvis, blood vessels, the lungs, the brain, the heart, abdomen and bones. CT is also often the preferred method of diagnosing many cancers, such as liver, lung and pancreatic cancers. During a CT scan, you lie still on a table. The table slowly passes through the center of a large X-ray machine. The test is painless. During some tests you receive a contrast dye, which makes parts of your body show up better in the image.

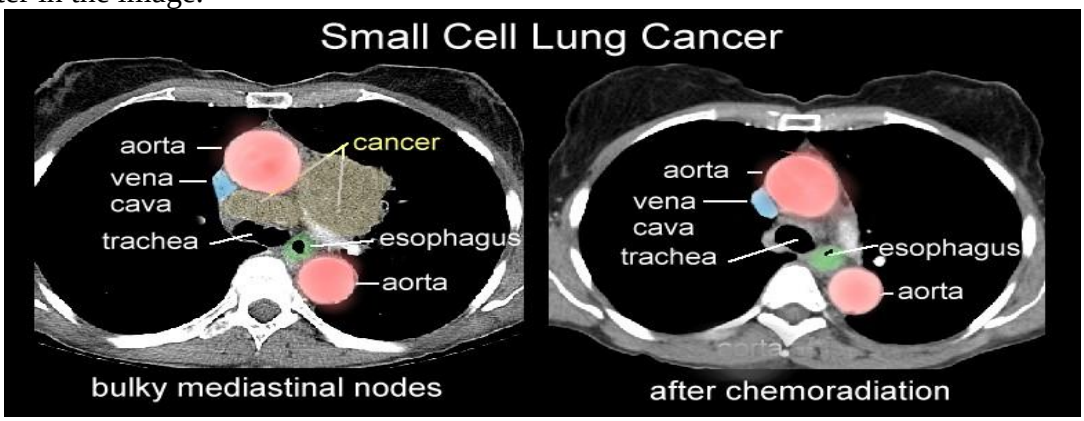

Figure 2: CT Image of Lung showing lung with cancer before treatment and healthy lung after treatment

CT is often used to evaluate:

i) Presence, size and location of tumors
ii) Organs in the pelvis, chest and abdomen
iii) Colon health (CT colongraphy)
iv) Vascular condition/blood flow
v) Pulmonary embolism (CT angiography)
vi) Abdominal aortic aneurysms (CT angiography)

vii) Bone injuries
viii) Cardiac tissue
ix) Traumatic injuries
x) Cardiovascular disease

1.3 Medical Resonance Imaging (MRI) Images

Magnetic Resonance Imaging (MRI) is a medical imaging technology that uses radio waves and a magnetic field to create detailed images of organs and tissues. MRI has proven to be highly effective in diagnosing a number of conditions by showing the difference between normal and diseased soft tissues of the body.

MRI is often used to evaluate:

i) Blood vessels
ii) Abnormal tissues
iii) Breasts
iv) Bones and joints
v) Organs in the pelvis, chest and abdomen (heart, liver, kidney, spleen)
vi) Spinal injuries
vii) Tendon and ligament tears

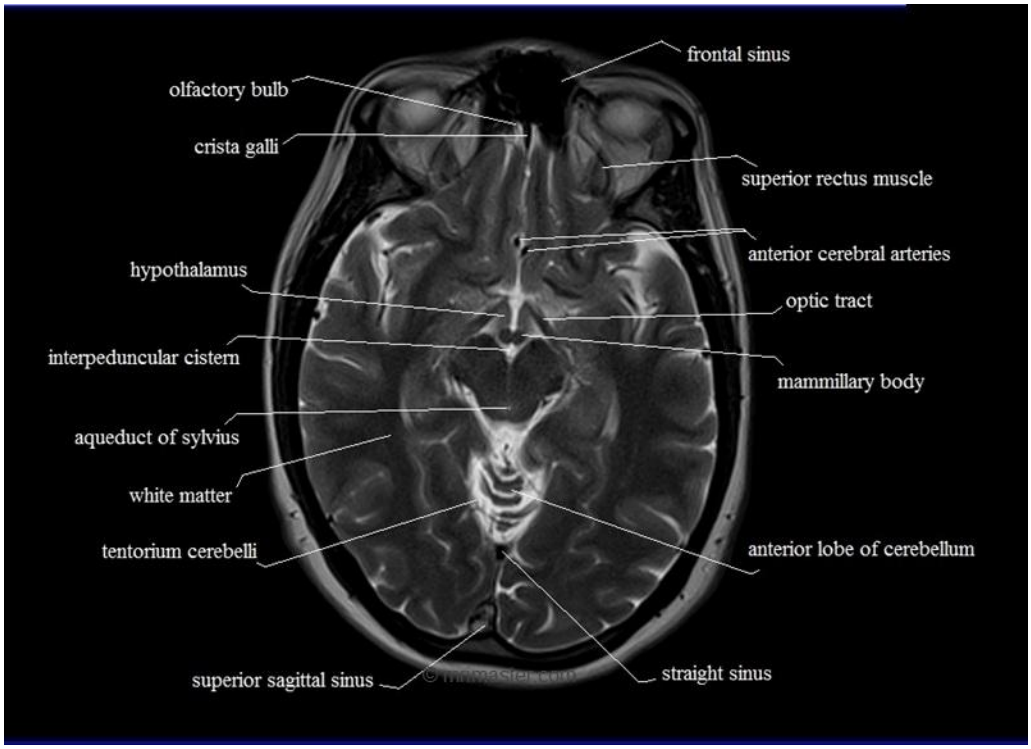

Figure 3: MRI Brain Anatomy

There are several other types of bio-medical imaging techniques such as X-rays, positron emission tomography (PET), PET-CT etc. The images taken through bio-medical imaging technique are normally taken for critical diseases and need an expert eye to examine the disease. Patients generally can't reach the specialist doctor either due to their own negligence or due to unavailability of one. In both cases the disease of the patient can become more and more critical and could lead to irreparable damage to the patients. There are several bio-medical image processing techniques which can help in such cases. One better and effective solution to the problem is Content-based image retrieval (CBIR) systems. CBIR techniques have proved themselves very useful in early detection of critical

diseases by making use of bio-medical images acquired from different normal and severe patients which are very useful data for the analysis of many hazardous diseases. These images are formed in image database and are used to analyse using CBIR techniques. The CBIR process comprises mainly of two stages. One is feature extraction and second is query matching. In feature extraction features such as colour, shape and texture of image are extracted and a feature vector is formed. Feature vector is an optimized form of an image. The second process in CBIR is query matching based on similarity measurement which uses the distance of the query from each image in database to find the closest image. The CBIR system developed can be very useful to detect the disease at early stage even at remote places where general clinical services are available. The detection results can be stored and the progression of disease can be monitored regularly. The patient's history can be saved from time to time in order to make it available easily to the specialist doctor so that he can diagnose and treat the patient efficiently.

2. Motivation

There has been a rapid growth in the field of medical science and in the methods of diagnosis. Medical Science, working hand in hand with engineering science created wonders. CBIR when used with medical imaging can be very useful to facilitate the doctors in treating the patient. Figure 4 explains the way in which CBIR can be used for Medical Images and how it can be preserved for future reference.

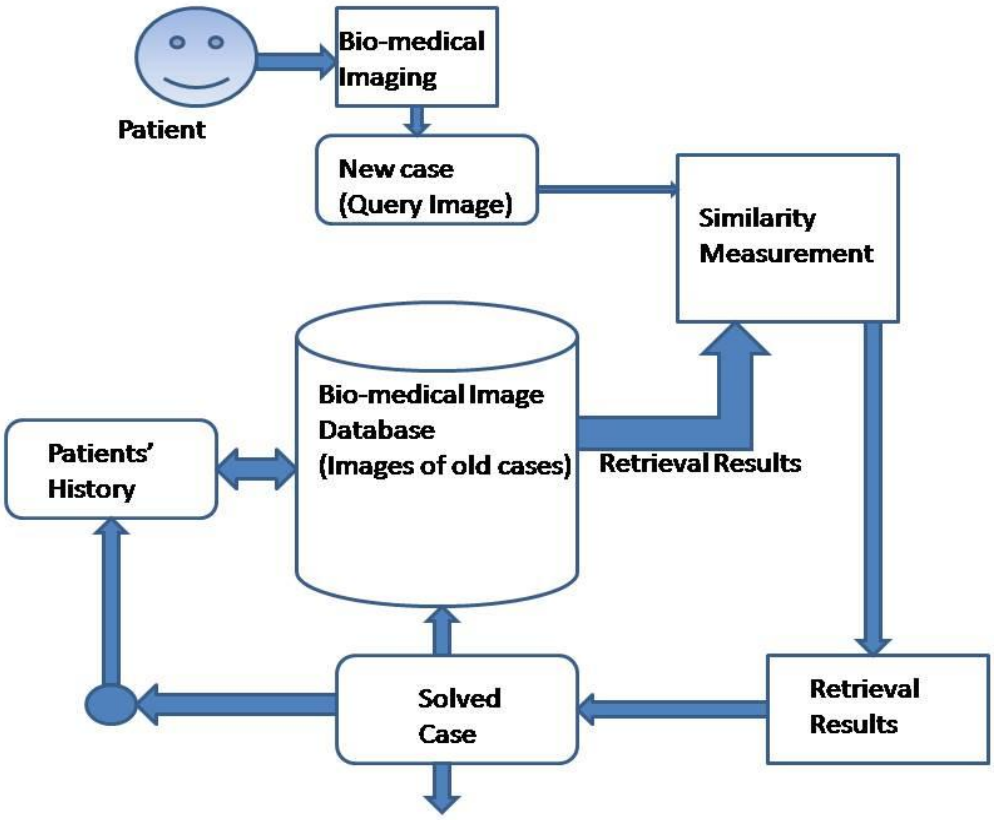

Figure 4: Use of CBIR in Bio-medical Imaging

Nowadays many hospitals are practising the use of medical imaging to diagnose, detect and record the patients' history. Many techniques of medical imaging are available

presently such as MRI, CT scans, X-rays, fundus imaging etc. With DICOM (Digital Imaging and Communications in Medicine) and advanced fundus imaging techniques, a standard for image communication has been set and patient information can be stored with the actual image, although still a few problems prevail with respect to the standardization. Many hospitals at present have enormous data of patients available at their records and to handle this much data by human eyes (such images need an expert eye) is not feasible. The characteristics of medical images differ substantially as discussed below:

- Noncolor: In medical imaging modalities, color is less relevant. For instance CT, MRI images, all form gray-scale image. Hence color cannot be used to describe images appropriately.
- Structure: The pathology that is visualised in medical images may result from only subtle or small alterations in the image pattern, if history of images is available for comparison at all. Usually intra individual differences in size, shape, and appearance of biological tissue are more prominent than diagnostically relevant alterations.
- Multidimensionality: Medical images are often 3D and hence more complex features must be used in the medical domain.
- Multimodality: Medical diagnostics frequently is based on more than only one set of image data, and imaging data is usually obtained from different modalities. The different nature of medical modalities does not support representing all such images using the same numerical features.
- Meta information: Medical images are only a part of the information that is relevant in a medical case. Gender, age and other data are required to compare diagnostics in medical cases.

So we need to develop some fully automated technique which can deal with such huge data and give us exact output that we want. Content-based image retrieval technique is best suited for such problem. In several articles, content–based access to medical images for supporting clinical decision–making has been proposed that would ease the management of clinical data and scenarios for the integration of content-based access methods into Picture Archiving and Communication Systems (PACS) have been created.

Further in future, CBIR technique can be used in combination with cloud computing and telemedicine technique. We can store large image databases of different diseases on cloud where we can deploy our automated CBIR designs and through telemedicine specialist doctors can diagnose the patient from distance. This advancement can take medical science to a different level.

3. Literature Survey

Zhao et al. [1] have used Gabor texture features for the analysis of liver CT images. These features extracted are used as feature vector for content based image retrieval. Mathieu et al. [2] have proposed a method in which they used model of wavelet coefficient distribution. They built signatures to characterize the image using distribution of coefficients of wavelet transform. The image retrieval is done by computing signature distances between query image and database image. Kenneth et al. [3] proposed a method to retrieve retinal images from a retinal image database which used an index vector to estimate the posterior probability of each of the categorized disease state. G. Quellec et al. [4] optimized wavelet transform for content based image retrieval. They used generic

indexing for images in database then they applied wavelet transform generate a signature for each image. They assigned weights to those signatures calculated and carried out optimization process of wavelet by finding best weights. For image retrieval they found out distance between signatures of query image and database image. Javed et al. [5] used scaled invariant feature transform (SIFT) and bag-of-words (BoW) model together to describe and differentiate 3D images of computed tomographic colonography computer-aided detection (CTCCAD).A Euclidean distance metric of theBoW histograms determines the similarity between thequery image and an image in the database. Murala et al. [6] proposed directional binary wavelet patterns for Bio-medical Image indexing and retrieval. They divided 8-bit grayscale image into 8 binary bit-planes and then applied binary wavelet transform (BWT) to extract multi-resolution binary images. Then they extracted local binary features using local binary patterns. They applied this method on three databases, OASIS-MRI, NEMA-CT and PolyU-NIRFD face database. Baby et al. [8] have proposed a method for content based image retrieval using dual-tree complex wavelet transform along with generalized Gaussian model (GGD) and kulback-leibler divergence (KLD) measurement. They have worked on Messidor database which is the largest publically available database which contains 1200 eye fundus color images. Naguibet al. [9] proposed a method for content based image retrieval of diabetic macular edema Images. In which they divided the macula into three concentric regions of which texture discontinuities are used to represent lesions in retina. The distance measure gives higher weights to lesions closer to the fovea to reflect the severity of DME. Yanling Chi et al. [10] have proposed Content-based image retrieval of multiphase CT images for focal liver lesion characterization in which they used hybrid generative-discriminative focal liver lesions (FLLs) detection method and a non-rigid B-spline registration method for localizing FLL on multiphase CT scan. Then they extracted multiphase density and texture features which represents image in numeric form, later they compared with features of query FLL using *L1* norm based similarity scores. Murala et al. [11] have proposed a novel feature extraction algorithm known as Local ternary co-occurrence patterns (LTCoP) for Bio-medical Image retrieval. The LTCoP encodes the co-occurrence of similar ternary edges which are calculated based on the gray values of center pixel and its surrounding neighbours. They carried out their experiments with this encoding in combination with Gabor transform. Sivakamasundari et al. [13] proposed a method using kirsch template based edge detection for content based retinal image retrieval system. They obtained structure and texture based features using blood vessel segmentation. For similarity matching of the query image with the database image they used Euclidian distance method. Jinsa et al. [14] proposed a method for content based image retrieval of CT images of lungs. They used Gray level co-occurrence matrix (GLCM) parameters and statistical parameters as feature vectors for retrieval process. Out of 22 GLCM parameters, the relevant 10 parameters are selected as artificial neural network classifier. Skewness is the statistical parameter selected by the classifier for retrieval. They used modified Bray-Curtis distance for similarity measurement. Romero et al. [15] proposed a new method for detection of microaneurysms. First they reduced non uniform illumination to normalize image grayscale content to improve dependence of images. After that they applied bottom-hat transform to remove reddish regions. Then they applied hit-or-miss transform to remove blood vessels from ROIs. They extracted two features out of which one feature distinguishes microaneurysms (which are round shaped) from vessels(which are elongated shaped), second feature gives a count of how many times radon transform is evaluated at the set of discrete angle values $\{0^\circ, 1^\circ, 2^\circ, \ldots, 180^\circ\}$.

4. Identified Open Research Problems

Bio-medical Image analysis is a very fast evolving area for researchers and those who have completed their results are converting their research into fully automated systems which are used for clinical use. Most of the open research is going on in following areas:-

1) Nowadays different Bio-medical image datasets are available for research as many hospitals are also diving into these research areas in collaboration with different research institutes to develop systematic approach to feature extraction and retrieval system developments.
2) We can take efforts in a focused way to develop optimized algorithms for Bio-medical Image analysis which could lead to design advanced diagnosis system which will enable medical experts to reach to the remote place patients and give their expert service from a distance.
3) Development of fast and accurate detection and retrieval algorithms which would help the doctors in diagnosing the severe diseases quickly and easily.
4) Portable, Cost-Effective Bio-medical Imaging.

5. Problem Definition

The rapid growth of severe diseases such as lung cancer, brain tumour, diabetic retinopathy etc. in the world and their exposure to irreparable damage would lead to:-

1) Need of complex analysis of Bio-medical images which would need an experts eye, but increasing ratio of number of patients to number of specialist doctors which will be felt significantly worldwide, it would not be easy to get the patient data analysed by medical experts.
2) So the amount of data to be analysed and interpreted by general physicians will be very large and the analysing experts' population will be very less so the engineering intervention is must in near future.
3) In these scenarios there will be very much need of development of a fully automated system which could categorise, detect and diagnose the complex diseases.
4) So there will be demand of automated texture analysis, correlation study and content based retrieval systems to handle such a large amount of data.

6. Future research plan

1) To developed a completely automated feature extraction method to assist in CBIR for Bio-medical Images.
2) Collection of different feature extraction methods to carry out CBIR for Bio-medical images.
3) Weight optimization of different feature extraction methods and use combination of different optimized feature extraction methods in CBIR for Bio-medical images.

7. Proposed work plan

Sr. No.	Phase	Details of work to be carried out
1	I	Literature Survey, Course work completion and implementation of existing algorithms, problem formulation.
2	II	Collection and classification of medical data and apply existing algorithms on the collected datasets and check performance. Modification in the existing techniques for improving performance.
3	III	Development of new feature extraction techniques for CBIR so that accurate retrieval and diagnosis of medical images take place.
4	IV	Work on weight optimization of combination of different feature extraction techniques that will be applied for CBIR.

References

[1] C. G. Zhao, H. Y. Cheng, Y. L. Huo, T. G. Zhuang," Liver CT-Image Retrieval Based on Gabor Texture", *Annual International Conference of the IEEE EMBS San Francisco, CA, USA* September 1-5, 2004 pp 1491-1494.

[2] Mathieu Lamard, G Cazuguel, G Quellec, L Bekri, C Roux, B Cochener, "Content based image retrieval based on wavelet transform coefficients distribution", *Annual International Conference of the IEEE EMBS* Aug 23-26 2007 pp 4532-4535.

[3] Kenneth W Tobin, M D Abramoff, E Chaum, L Giancardo, V. PriyaGovindasamy, T P Karnowski, M T S Tennant, S Swainson, "Using a patient image archive to dignose retinopathy" ,*Annual International Conference of the IEEE EMBS* Aug 20-24 2008 pp 5441-5444.

[4] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, C. Roux, "Wavelet optimization for content based image retrieval in medical databases" *Medical Image Analysis* 14(2010) pp 227-241.

[5] Javed M. Aman, Jianhua Yao, Ronald M. Summers," content-based image retrieval on ct colonography using rotation and scale invariant features and bag-of-words model", *IEEE ISBI* 2010 pp-1357-1360.

[6] Subrahmanyam Murala, R. P. Maheshwari, R. Balasubramanian," Directional Binary Wavelet Patterns for Biomedical Image Indexing and Retrieval" *Springer J Med Syst* (2012) 36 pp 2865–2879.

[7] I. Lazar and A. Hajdu," Retinal Microaneurysm detection through local rotating cross-section profile analysis", *IEEE transactions on medical imaging.* 32(2), February 2013 pp 400-407.

[8] C.G. Baby, D. A. Chandy," Content-Based retinal image retrieval using dual-tree complex wavelet transform", *International conference on Signal processing, Image processing and Pattern Recognition*- 2013 pp 1-5.

[9] Aya M Naguib, A M Ghanem, A S Fahmy, "Content based image retrieval of diabetic macular edema images", *IEEE CBMS* 2013 pp 560-562.

[10] Yanling Chi, Jiayin Zhou, Sudhakar K. Venkatesh, Qi Tian and Jimin Liu," Content-based image retrieval of multiphase CT images for focal liver lesion characterization", *Medical Physics, Vol. 40, No. 10*, October 2013, pp (103502)1-13.

[11] SubrahmanyamMurala,Q.M.JonathanWu," Local ternary co-occurrence patterns a new feature descriptor for MRI and CT Image retrieval.", *Neurocomputing 119* 2013 pp 399-412.

[12] M. N. Ashraf, Z. Habib, M. Hussain," Texture feature analysis of digital fundus images for early detection of diabetic retinopathy", *International conference on computer graphics, imaging and visualization* 2014 pp 57-62.

[13] Sivakamasundari J, Kavitha G. Natarajan V, Ramakrishna S, "Proposal of content based retinal image retrieval system using kirsch template based edge detection", *IEEE 3rd International conference on informatics, electronics & vision* 2014 pp 607-611.

[14] Jinsa Kuruvilla and K. Gunavathi," Content Based Image Retrieval for CT Images of Lungs", *International Journal of Soft Computing 9 (6)* 2014 pp 386-390.

[15] R Rosas-Romero, J Martinez-Carballido, J Hernandez-Capistran, L J Uribe-Valencia, "A method to assist in the diagnosis of early diabetic retinopathy: image processing applied to detection of microaneurysms in fundus images", *computerized medical imaging and graphics 44* 2015 pp 41-53.

Mr. Galshetwar Gajanan Madhavrao
(Research Scholar)

Dr. L. M. Waghmare
(Supervisor)